Hydrology Research

© 2022 The Authors

Hydrology Research Vol 53 No 9, 1150 doi: 10.2166/nh.2022.003

How do we identify flash droughts? A case study in Central European Croplands

Pedro Henrique Lima Alencar [a,b,*] and Eva Nora Paton [a]

[a] Ecohydrology and Landscape Evaluation, Institute of Ecology, Technical University of Berlin, Berlin, Germany
[b] Agricultural Engineering, Agricultural Sciences, Federal University of Ceará, Benfica, Fortaleza, Brazil
*Corresponding author. E-mail: pedro.alencar@campus.tu-berlin.de

PHLA, 0000-0001-6221-8580; ENP, 0000-0002-5619-9958

ABSTRACT

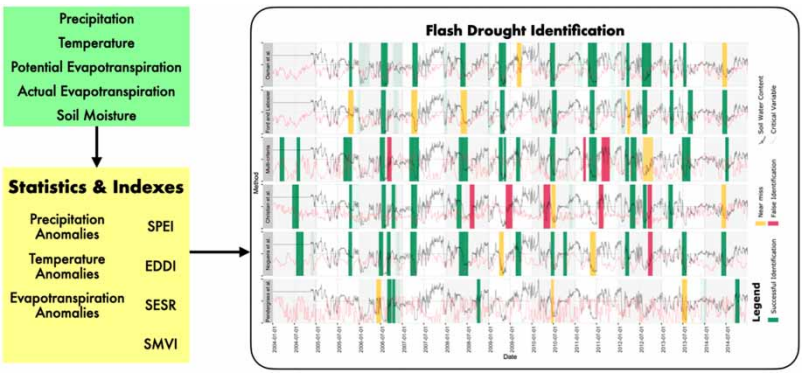
Many definitions and delineation methods exist for identifying flash droughts (FDs), which are events of rapid and unusual large depletion of root-zone soil moisture, in comparison to average moisture conditions, due to climatic compound conditions over a short period of several weeks. Six FD identification methods were compared to analyse their functioning using data from several experimental cropland sites across Central Europe. Co- and misidentification of the FD time series were assessed using confusion and synchronicity metrics on a local scale. Even though a large degree of synchronicity of individual FD events was observed, some divergence in drought periods was detected, which was related to four intrinsic differences in the underlying FD definitions: (1) type of critical variable; (2) velocity of drought intensification; (3) pre-set threshold values for final depletion and/or (4) minimum length of the duration of FDs. To balance the strengths and weaknesses of those methods that are not based on soil moisture, we suggest using an ensemble approach for event identification, which is validated in this study for the temperate central European region. In doing so, the current unclearly defined sub-types of FDs can be detected, regardless of the different combinations of compound drivers and differences in intensification dynamics. All methods were implemented in an R package and are available as a Shiny app for the public.

Key words: climatic compound events, confusion matrix, event identification, flash drought, synchronicity metrics

HIGHLIGHTS

- The multiple methods proposed to identify flash droughts (FDs) show substantial disagreement.
- Soil moisture is the key variable to identify FDs in croplands; however, such data are scarce, and a method based on proxy variables is necessary.
- A multi-index or multi-method should be favoured in identifying FDs, as a single proxy (to soil moisture) may cause significant misidentification.

GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

1. INTRODUCTION

Droughts are among the most extreme climatic weather events that threaten food security (FAO 2021). They have negative impacts on the global food–energy–water nexus and the sustainable development goals (D'Odorico *et al.* 2018). Droughts are generally characterised by unusually high levels of rainfall deficit, runoff deficit or soil moisture deficit (Palmer 1965; Mishra & Singh 2010) and are expected to increase in many regions of the world in terms of frequency, severity and duration under current and future climate change conditions (Lesk *et al.* 2016; Samaniego *et al.* 2018).

Flash droughts (FDs) are a special form of drought. In contrast to the descriptions of classical droughts (de Araújo & Bronstert 2015; Oikonomou *et al.* 2020), FDs are characterised by rapid onset and relatively short durations (Otkin *et al.* 2018; Lisonbee *et al.* 2021). They are associated with severe and immediate soil moisture depletions, resulting in plant water stress and mortality (Ford & Labosier 2017; Liang & Yuan 2021; Osman *et al.* 2021).

One of the first FD studies was by Peters *et al.* (2002), who were among the first to study single short drought events in late summer, which were characterised by the concurrence of low antecedent moisture and unusually high temperature. Interest in FDs has increased over the last few years, motivated by extreme FD occurrences in the USA, Russia and China, which caused extreme impacts on managed vegetation, disruption to the global food supply and increased wildfires (Otkin *et al.* 2013; Mo & Lettenmaier 2016; Christian *et al*. 2019a, 2019b; Liang & Yuan 2021). An FD in Australia during the spring of 2019 is thought to have played a central role in the massive forest fires that consumed over 1.6 million hectares (Nguyen *et al.* 2021).

Over the last two decades, multiple methods have been proposed to identify FD events (Lisonbee *et al.* 2021), yet there is no consensus on what an FD entails and how they may be defined in terms of onset, duration, the velocity of intensification and absolute or relative changes (Osman *et al.* 2021). In their recent literature review, Lisonbee *et al.* (2021) identified as many as 20 studies with different definitions using climate variables or indexes related to soil moisture, air temperature, precipitation, and actual and potential evapotranspiration (ET). Eleven of these definitions included an interval of intensification or rapid onset as part of flash flood delineation, whereas nine studies merely considered short-term drought events as FDs. A method comparison study by Osman *et al.* (2021) identified four core types where the definition of a heatwave FD (Mo & Lettenmaier 2015) was based on temperature anomalies, rapid soil drying (Ford & Labosier 2017; Yuan *et al.* 2019), actual and/or potential ET anomalies (Christian *et al.* 2019a, Pendergrass *et al.* 2020) and multi-criteria indexes (Chen *et al.* 2019). For the USA, they showed that the FD frequency, spatial extent and onset would vary significantly depending on which definition is used. They also suggested that a root-zone soil moisture-based method effectively captures the FD onset in both humid and semi-arid regions. Other than the study by Osman *et al.* (2021), which focused on the identification of spatial differences in delineated FDs, there have been no systematic studies on the temporal divergence and synchronicity of delineated FDs. At the same time, little is known about FD dynamics in Central Europe and it is not known which FD method would apply to this region, given that the present methods have so far been used for FD identification mainly outside Europe. Additionally, little efforts have been made towards an ensemble method for FD identification. In this study, we propose a method of this nature tested for areas in Central Europe.

Due to the lack of a general definition, we define an FD as the process of rapid, accelerated and unusually large depletion of root-zone soil moisture, in comparison with 'average' moisture conditions, due to the simultaneous or concurrent occurrence of two or more atmospheric and/or weather conditions over a short period of several weeks during the main growing season.

The objective of this study was to compare, in a local/point (climatological station) scale, the functioning of six recently developed FD identification methods with data from four well-monitored experimental cropland sites in Central Europe, by assessing co- and misidentification of FD time series using similarity and synchronicity metrics. We selected two soil moisture-based methods (Ford & Labosier 2017; Osman *et al.* 2021) and four indirect methods that used single or multiple climatic variables or indices for FD delineation (Christian *et al.* 2020; Noguera *et al.* 2020; Pendergrass *et al.* 2020, and a multi-criteria method proposed by the authors and validated for Central Europe). The methods were implemented in an R package and a Shiny app available to the public.

2. MATERIALS AND METHODS

2.1. FD identification methods

The following six FD identification methods were selected on the basis that they used station data as input and, following our definition, included a clear definition of the rapid onset of water limitation. The first two methods are based on soil moisture

and the other four used indirect proxies of drought conditions, such as anomalies of rainfall, temperature and the ratio of actual and potential ET. The methods used are described as follows:

M1: Osman *et al.* (2021) borrowed the concept of volatility from stock market analysis techniques for the analysis of rapid soil moisture changes. According to them, an FD occurs when the one-pentad (5 d) running average for root-zone soil moisture content falls below the four-pentad (20 d) running average for a period of at least four pentads, with soil moisture at the end of this period dropping below the 20th percentile for that time of year (Figure 1(a)).

M2: Ford & Labosier (2017) identified FDs as periods when the pentad-average 0–40 cm volumetric water content declines, from at least the 40th percentile to below the 20th percentile, in four pentads or less (Figure 1(b)).

M3: The multi-criteria method is a new method that uses a set of 10 anomalies and indexes derived from weekly precipitation, temperature and potential ET data. It calculates a score for each week equivalent to the proportion of indicators that meet or surpass the respective pre-set thresholds. Weeks with a score higher than 0.65 and an absolute change of score (Δscore) higher than 0.25 over up to 3 weeks are classified as FDs (Figure 1(c)). The event duration is computed as the time from the beginning of intensification until the score is below 0.65. A full description of this method is provided in the Supplementary Material.

M4: Christian *et al.* (2020) used the standardised evaporative stress ratio (SESR), which is derived as the *z*-score of the quotient of actual to potential ET rate values for a specific pentad. They used four criteria which FD events were required to have: (1) a minimum length of five SESR changes, equivalent to a length of six pentads (30d minimum length); (2) a final SESR value below the 20th percentile of SESR values; (3) SESR changes must be at or below the 40th percentile between individual pentads and no more than one SESR change above the 40th percentile following the previous criterion and (4) an overall mean change in the SESR during the entire length of the FD must be below the 25th percentile in the SESR (Figure 1(d)).

M5: Noguera *et al.* (2020) used the standard precipitation evapotranspiration index (SPEI) on a short timescale (1 month) and performed calculations based on a temporal frequency of 1 week (four per month). To identify the rapid onset of a drought event, the change in the SPEI for each week, in periods of 4 weeks, was calculated and the onset of an FD was defined as involving a change in the SPEI equal to or less than -2 SPEI units (*z*-values) over an intensification period of 4 weeks. Further, final SPEI values had to be equal to or less than -1.28 SPEI units (Figure 1(e)).

M6: Pendergrass *et al.* (2020) utilised the evaporative demand drought index (EDDI) following Hobbins *et al.* (2016) and Lukas *et al.* (2017), which is calculated based only on the potential ET using the Penman–Monteith equation. The method identifies an FD when a 50% increase in the EDDI over 2 weeks is sustained for at least another 2 weeks (Figure 1(f)).

Workflow and key characteristics are summarised in Figure 1 and Table 1. Readers interested in a more detailed description of the percentiles and thresholds are referred to the original papers.

The six methods used one or more climate variables (rainfall, temperature, soil moisture and actual and potential ET). Further, they all share an underlying set of characteristics:

1. FDs evolve rapidly, with an intensification period lasting between 2 and 4 weeks.
2. The final conditions at the end of an FD lean towards extreme values, which are often characterised by the variable reaching values under the 20th percentile or, in some, a *z-score* value over ± 1.
3. FDs are considered seasonal phenomena and are identified based on the expected values of climatic variables for each specific time of the year subdivided either in pentads or weeks.
4. FDs depend on crossing certain thresholds and are thought to be correctly identified if, and only if, environmental conditions meet a set of predefined rules.

The key differences between the methods are how the FD variables and durations are defined. Methods M1 – Osman *et al.* and M2 – Ford & Labosier take a direct approach to assess plant water availability using soil water data, whereas all other methods use proxy variables, which are likely to be less accurate, while simultaneously overcoming severe data limitations. Additionally, differences exist in the definition of the onset of the FD periods, the time resolution used (weeks and pentads), the minimum period over which it should be sustained and the maximum duration beyond which it might be considered a 'normal' drought (Table 1).

Additionally, the definition of event duration varies among methods. Most methods (M1 – Osman *et al.*, M2 – Ford & Labosier, M3 – multi-criteria and M6 – Pendergrass *et al.*) assume that the duration of an FD event comprises the intensification phase (between 1 and 3 weeks), a period of persistence of the dry conditions (between 2 and 4 weeks) ending with the recuperation phase when the key variable (climatic or groundwater conditions) surpasses a threshold (Table 1). The methods of M4 – Christian *et al.* and M5 – Noguera *et al.* have different assumptions regarding duration. M4 – Christian *et al.* 2019a assume that the FD event is limited to the intensification period, with the FD event lasting only as long as there is a trend towards dry

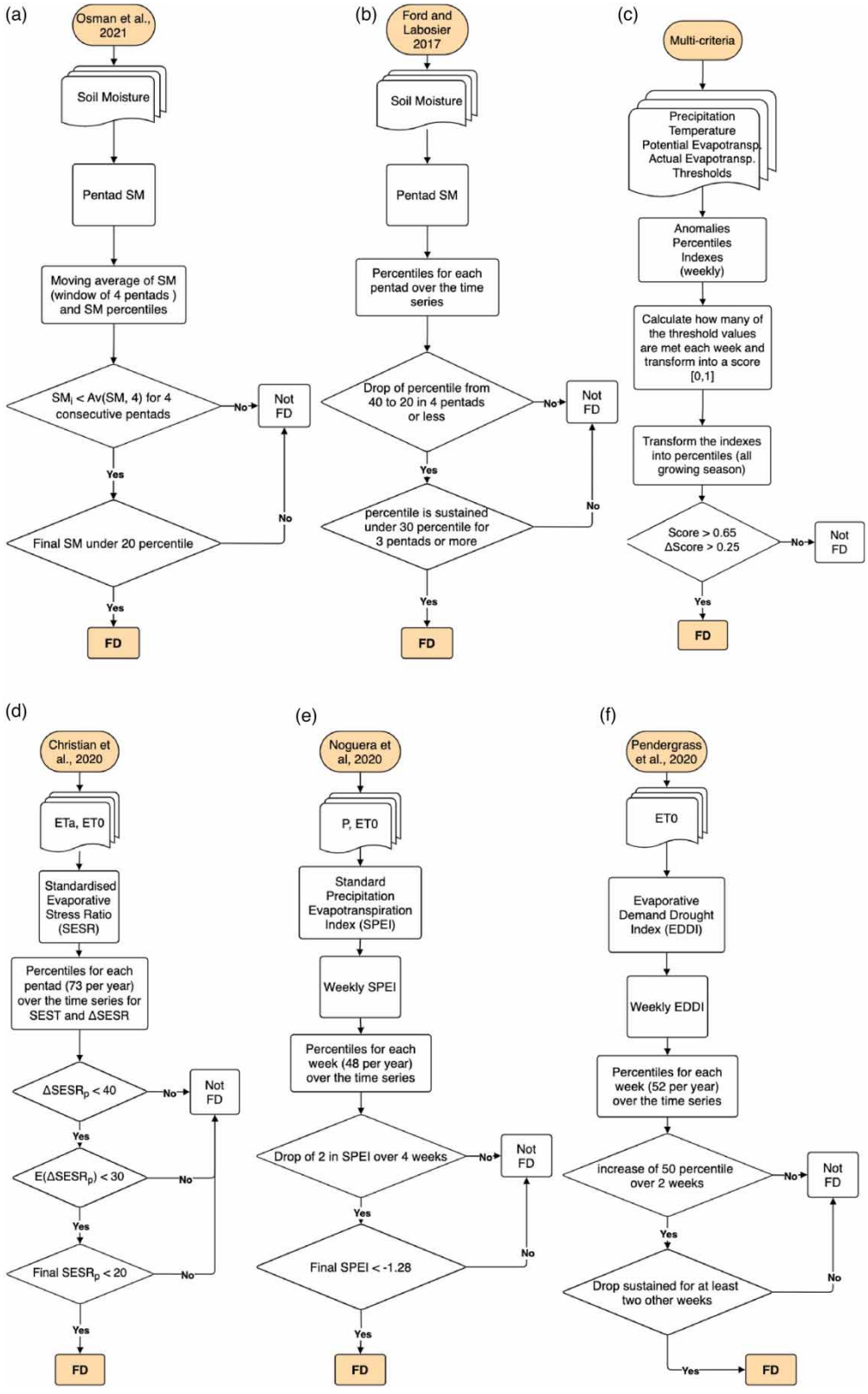


Figure 1 | Flowcharts of the six methods for FD identification: (a) M1: Osman *et al.* (2021); (b) M2: Ford & Labosier (2017); (c) M3: novel multi-criteria method; (d) M4: Christian *et al.* (2020); (e) M5: Noguera *et al.* (2020); and (f) M6: Pendergrass *et al.* (2020). The implementation of all methods is available in the supplements of this paper as an R package.

Table 1 | Comparison of key variables, statistics and threshold criteria for the detection of FDs in all six methods

	Method	Variables	Original dataset type[4]	Statistics	Onset rate	Total duration
M1	Osman *et al.* (2021)	Soil moisture	Grid (CONUS)	Soil moisture volatility index (SMVI) Running averages (RA)	4 weeks	Onset duration + while SM lower than 4-pentad running average
M2	Ford & Labosier (2017)	Soil moisture	Station (Eastern USA)	Percentiles	4 pentads	Onset duration + until SM>30th p.
M3	Multi-criteria	Precipitation temperature potential evapotranspiration	Station (Central Europe)	Anomalies indexes percentiles	4 weeks	Onset duration + until 2 consecutive weeks with score<0.65
M4	Christian *et al.* (2019a, 2019b, 2020)	Actual evapotranspiration potential evapotranspiration	Grid (Eastern USA[5] and SW Russia)	Standardised evaporative stress ratio (SESR)	6 pentads	Onset duration + until 2 consecutive weeks with an increasing SESR
M5	Noguera *et al.* (2020)	Precipitation potential evapotranspiration	Station (Spain)	Standard precipitation evapotranspiration index (SPEI)	4 weeks	After onset, until SPEI>−1.28
M6	Pendergrass *et al.* (2020)	Potential evapotranspiration	GRID (CONUS)	Evaporative demand drought index (EDDI)	2 weeks	Onset duration + until EDDI increases towards wet conditions

[4]The data type used in the original application/publication of the method. Grid indicates methods that were implemented with reanalysis or remote sensing data. Station indicates methods that were originally implemented with weather station (direct measure) data.
[5]The Christian *et al*. method studied multiple areas in the continental USA: the Great Plains, Corn Belt, and Great Lake regions, in the states of Georgia, Kansas, Iowa, and Minnesota.

conditions and not having a threshold for the recuperation phase (Christian *et al.* 2019a). M5 – Noguera *et al.*, although characterizing all three phases (intensification, persistence and recuperation), decided in their work not to consider the intensification phase as part of the FD duration. In this study, the application of the method of Noguera *et al.* (2020) includes the intensification period in the computation of the total duration of FD events.

All six methods were implemented in R programming language and are organised in an R package named 'fdClassify'[1] and in a Shiny App named 'FD-Viz'[2].

2.2. Comparison metrics

Two types of metrics were used to compare the six FD methods: synchronicity metrics and the confusion matrix. The method of Osman *et al.* (2021) was used as a reference method as it was considered the method that most closely followed our definition of an FD as given in Section 1 (using the rapid soil moisture decline as a key variable for FD identification). This method was evaluated against measured soil moisture data (see Section 2.3) in the root zone and, therefore, appeared to be particularly suited to reproduce the FD dynamics of croplands.

2.2.1. Synchronicity metrics

The concept of synchronicity based on the work of Kemter *et al.* (2020) was employed to compare the rate of identification of FD events and the intervals, which were correctly identified as intervals with no FD, for a weekly resolution. Synchronicity metrics were originally developed to analyse whether extreme floods occur concurrently with the same timing in larger basins (Kemter *et al.* 2020). It is based on the two synchronicity metrics $sync_1$ and $sync_0$, which are defined as the average proportion of the successful identification of FD events and no FD events, respectively. Here, the Osman reference method was compared separately with the other methods. The metrics can take values between zero and one; therefore, a perfect

[1] Link to fdClasify git repository: https://github.com/pedroalencar1/fdClassify.
[2] Link to Flash Drought Visualization tool (FD-Viz): https://pedroalencar.shinyapps.io/FD-Viz/.

agreement between one method and the Osman *et al.* (2021) reference method occurs if both metrics are equal to one. If only $sync_1$ is close to one and $sync_0$ is close to zero, it would mean that all FD events were identified as in the reference method, but the interval times without FDs were all identified incorrectly (by identifying many more FD events, where the reference method did not find one). The two metrics $sync_1$ and $sync_0$ are further summarised by their harmonic mean $\overline{sync}$ (which is given by the reciprocal of the arithmetic mean of their reciprocals) to represent the trade-offs between the two synchronicities.

2.2.2. Confusion matrix

A confusion matrix is a tool borrowed from data science (James *et al.* 2013), from which several metrics can be derived to evaluate the rate of true and false identifications of an event using a weekly resolution. Using the Osman *et al.* method as the reference method, for all weeks the matrix assesses the following scores for each FD identification method individually (Figure 2 presents an example of the comparison of M1 with M2):

- a true positive (TP) score for a true identification of an FD: both methods identified an FD;
- a true negative (TN) score for a correct non-identification: both methods did not identify an FD;
- a false positive score (FP) for a wrong identification: the Osman *et al.* method did not identify an FD but the other method did and
- a false negative score (FN) for an incorrect identification: the Osman *et al.* method did identify an FD but the other method did not.

The four scores are then summarised into several confusion metrics that describe the degree of similarity (equations in Table 2):

1. The true positive rate (TPR) and positive prediction value (PPV) evaluate the performance of the tested method to correctly replicate an FD.
2. The true negative rate (TNR) and negative prediction value (NPV) evaluate the performance of the tested method to correctly replicate the intervals between FDs.
3. The Matthews correlation coefficient (MCC) summarises resemblance and considers imbalanced datasets for which one class of events, in this case, the FDs, is much smaller than the other class, that is, no FD periods (Matthews 1975; Delgado & Tibau 2019; Chicco *et al.* 2021).

The metrics range from 0 to 1, with values close to 1 indicating a high degree of similarity and values close to 0 indicating little resemblance.

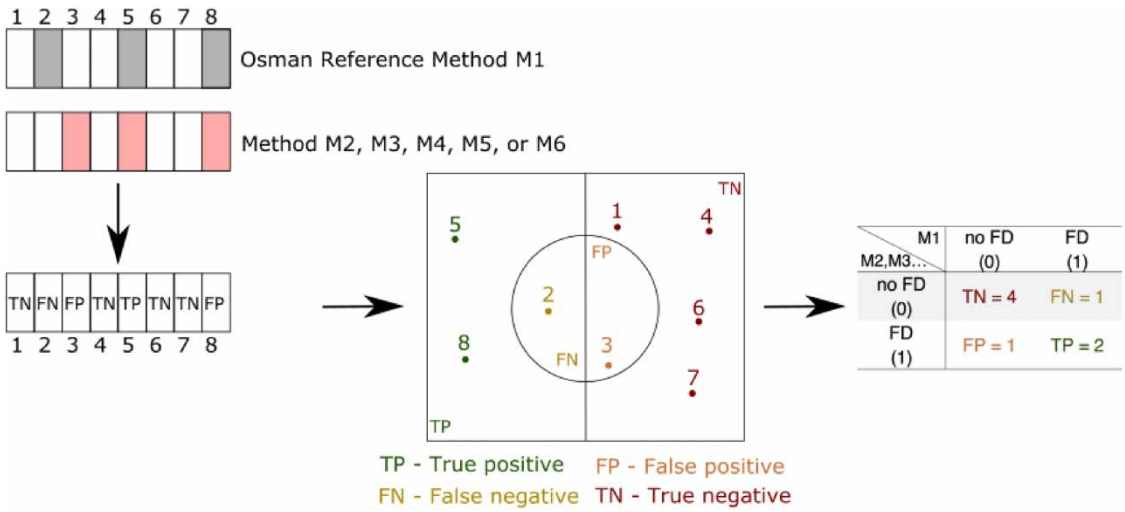


Figure 2 | Graphic representation of the confusion matrix for an example dataset of 8 weeks. The M1 – Osman reference method is compared with method M2 – Ford (or any other method) for TN, TP, FN and FP identification of FDs. Grey intervals: a FD occurred according to M1; red intervals: a flash drought occurred according to M2, white interval: no FD occurred according to either method. The plot in the middle visualises the same identifications (TP and FN) on the left side and the right side shows opposing identifications (FP and TN). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2022.003.

Table 2 | Definition of metrics from the confusion matrix

Metrics	Equations
True negative rate (TNR)	$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
Negative predictive value (NPV)	$\mathrm{FNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}$
True positive rate (TPR)	$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
Positive predictive value (PPV)	$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
Matthews correlation coefficient (MCC)	$\frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$

2.3. Data and study areas

The climate series from cropland stations of the FLUXNET2015 dataset (Pastorello *et al.* 2020) was used for FD identification, as this dataset provided high-quality measured data for most of the required climate variables. Daily station data of the variables of precipitation, temperature, soil water content, latent and sensitive heat fluxes, wind speed and relative humidity, with durations ranging between 11 and 14 years, were available for four stations in Central Europe, where soil water data were the most limiting criteria for station selection. Table 3 contains information on soil type and location.

Potential ET was calculated using the standard method described by Allen (1998). The actual ET was calculated using Equation (1), which relates evaporation to latent heat and temperature as follows:

$$\mathrm{ET} = \frac{H_l}{l_v \rho_w} \tag{1}$$

where H_l is the latent heat flux, ρ_w is the density of water and l_v is the latent heat of vaporisation, which, in turn, is a function of air temperature calculated following the procedures of Rogers (1989).

3. RESULTS AND DISCUSSION

3.1. Station-by-station comparison

FDs identified by all six methods for the four stations showed many commonalities as well as disparities (Figure 3(a)–3(d)[3]). Each time series plot in Figure 3 shows the daily soil moisture content in a black line and the critical variable of each method

Table 3 | Experimental cropland stations with information on location, duration of available soil moisture series, soil type and climate zone

Station[a]	Duration (years)	Latitude	Longitude	Elevation (m.a.s.l.)	Temperature (°C)	Precipitation (mm)	Climate[c]	Soil	Reference (doi)
BE-Lon	11	50.55	4.75	167	11.4	766	Cfb	Luvisol (*silt loam*)	10.18140/FLX/1440129
DE-Geb	14	51.10	10.91	161.5	9.7	531	Cfb	Chernozem (*clay loam*)	10.18140/FLX/1440146
DE-Kli	11	50.89	13.52	478	7.8	811	Cfb	Podsol (*silty loam*)	10.18140/FLX/1440149
IT-BCi	11	40.52	14.96	20	17.9	1199	Csa	(–)[b]	10.18140/FLX/1440166

[a]Data source: fluxnet.org/data/fluxnet2015-dataset/.
[b]Data not available.
[c]Climate classification according to the Köppen classification system.

[3] For detailed visualization of the data in Figure 3, please visit the Shiny App (FD-Viz): https://pedroalencar.shinyapps.io/FD-Viz/.

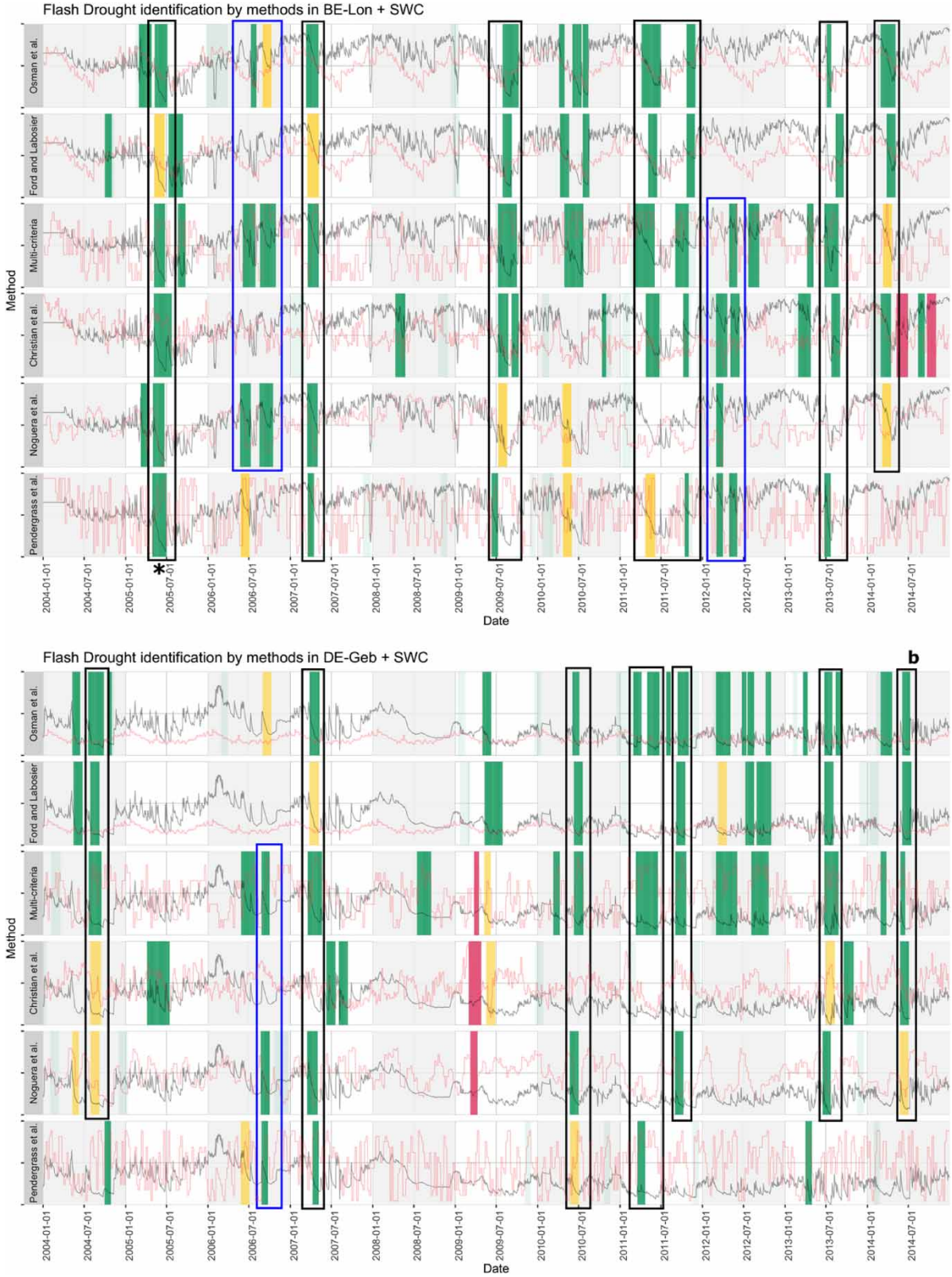


Figure 3 | FD events for six methods and four stations (a: BE-Lon, b: DE-Geb, c: DE-Kin and d: IT-BCi) for the time period 2004–2014. Colour coding: green bars show the start and end of FDs, yellow bars show near-misses and red bars show clear false identifications. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2022.003. (continued).

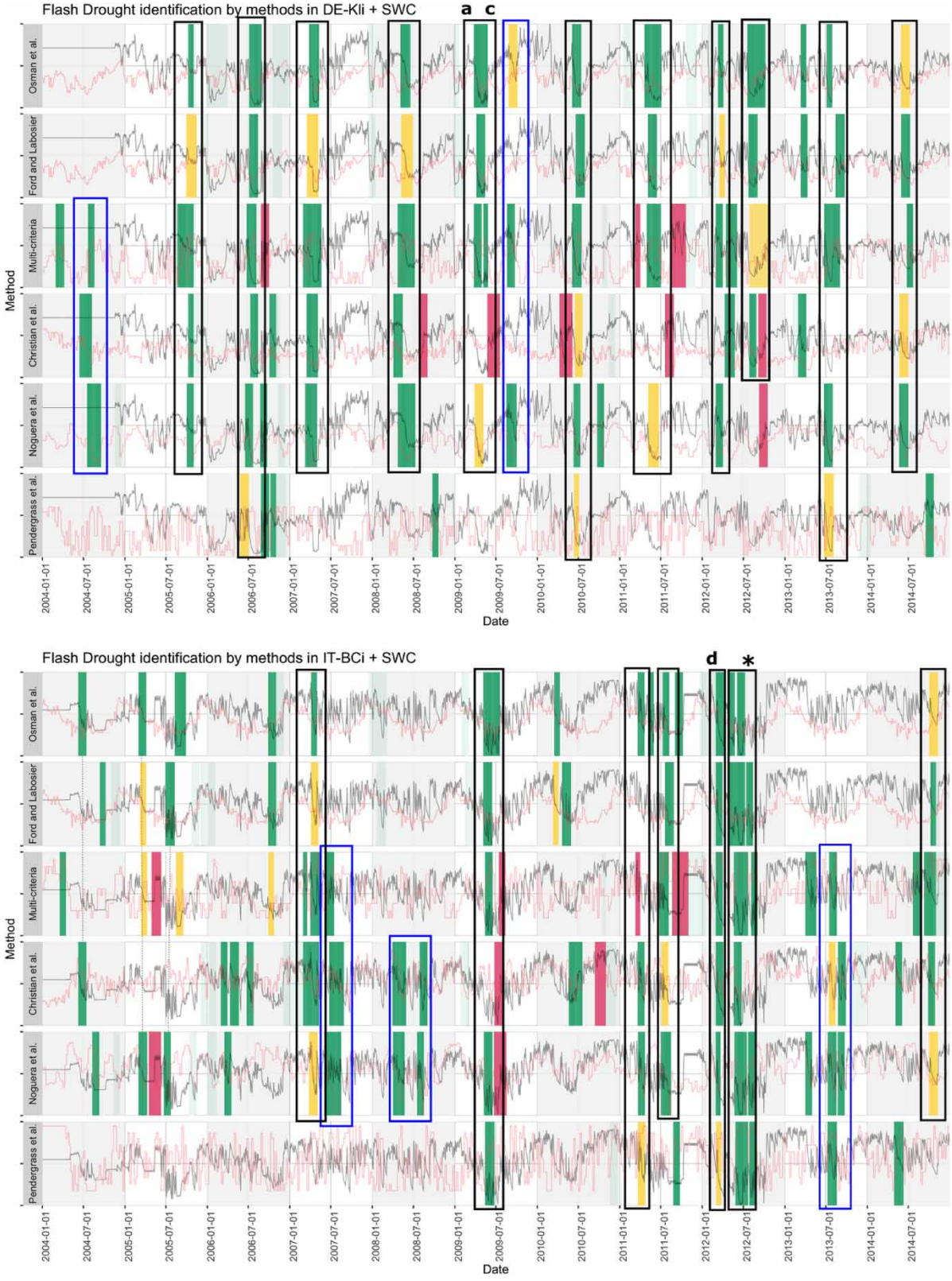


Figure 3 | (Continued).

in red: the 20th percentile for M1 – Osman *et al.* and M2 – Ford & Labosier, the summary score for M3 – the multi-criteria method, the SESR for M4 – Christian *et al.*, the SPEI for M5 – Noguera *et al.* and the EDDI for M6 – Pendergrass *et al.* The green bars mark the duration of identified FDs and the yellow bars indicate periods in which a method missed identification by a small fraction of its respective score or threshold.

In the majority of the 11-year study period, all methods detected at least one FD per year, except in 2005 and 2008. The method M6 – Pendergrass *et al.* yielded the lowest number of events at all stations (4–7), whereas the other methods identified considerably more events varying between 8–10 (M2), 13–14 (M3), 5–15 (M4) and 5–15 (M5). M1 – Osman *et al.*, with 11–18 events, identified the largest number of events.

Concurrent FD identification of five methods, marked with black frames (three methods, marked with blue frames), occurred 6 (2) times at Be-Lon, 10 (2) times at DE-Kli, 7 (1) times at DE-Geb and 6 (3) times at IT-Bci, with some small variations in the exact start time and duration. Station DE-Geb showed the overall best overlap for all methods, followed by Be-Lon. De-Geb and IT-Bci were characterised by longer periods, with no clear overlap between 2009 and 2012 for De-Geb. There was also no clear overlap between 2004, 2006 and 2010 for IT-Bci.

The two soil moisture-based methods, M1 and M2, showed the highest similarities overall. Every single event of M1 as the reference method was assessed for consistency with our FD definition. Other events that were not detected by the reference method were also evaluated to determine if they should have been delineated. It is noteworthy to mention that there were no concurrent FD identifications that did not include the reference method M1.

Concurrent identification among multiple methods would considerably increase if the *just-missed* events (yellow coding in Figure 3) were also considered, suggesting that some of the thresholds might have been set too rigidly. The red bars in Figure 3 mark obvious false identifications due to artefacts of the percentile approaches. This signifies some threshold crossing, but at the same time, the corresponding time series does not show any rapid decrease in absolute terms. Methods M3–M5 showed 24 false events altogether, with most of them in the early spring or autumn seasons. Instances of false identification may be reduced if additional rules regarding absolute changes in climate variables are introduced into the methods.

3.2. Method comparison using synchronicity metrics

The metrics of synchronicity $sync_1$ (two methods identifying the same FD periods) and $sync_0$ (two methods identifying the same no FD intervals) and their harmonic mean values $\overline{sync}$ are given in Figure 4 for methods M2–M6, in comparison to M1 as the reference method, for all four stations. None of the methods showed perfect agreement with the reference method, for which all three metrics would have equalled 1.

Methods M2–M5 presented intermediate values for the harmonic mean $\overline{sync}$ indicating similar behaviour regarding FD and no drought identification, in comparison to the reference method M1, with $\overline{sync}$ values ranging between 0.47 (M2: Ford & Labosier) and 0.39 (M5: Noguera *et al.*). The multi-criteria method M3 showed the highest $sync_1$ values overall, particularly for the station BE-Lon, directly followed by M2. For methods M2–M5, the $sync_1$ value at the station DE-Geb was significantly smaller than that of the other stations, indicating that the co-identification of FDs in comparison with the method M1 was the smallest at this station.

The method M6 (Pendergrass *et al.*) showed the largest differences from the reference method, with a $\overline{sync}$ value of only 0.2. However, M6 had large values for $sync_0$ (0.9), thus showing a great agreement in not identifying periods of no FD. Moreover, the small value of $sync_1$ (0.1) clearly shows that M6 rarely co-identified an FD simultaneously with M1.

3.3. Method comparison using confusion matrix metrics

The metrics given by the confusion matrix in Figure 5 provide a slightly different picture compared to the synchronicity metrics. The MCC describing the general resemblance of co-identification between the reference method M1 – Osman *et al.* and the other methods was the highest for M2 – Ford & Labosier (the median value of 0.43), followed by the multi-criteria method (0.28). The other three methods had small values close to zero, indicating little overall resemblance.

All methods had high values for the TNR and the NPV, with M2 – Ford & Labosier and M6 – Pendergrass *et al.* having the highest TNR (above 0.9), and all NPV metrics varied between 0.7 (Pendergrass *et al.*) and 0.8 (Ford & Labosier). Hence, the methods did well in identifying intervals with no FD, relative to the reference method. However, there were significantly larger discrepancies for the TPR and PPV, M2 – Ford & Labosier had reasonably high values (0.75 and 0.43, respectively), directly followed by the multi-criteria method (0.55 and 0.51) and the method of Christian *et al.* (0.45 and 0.41), whereas

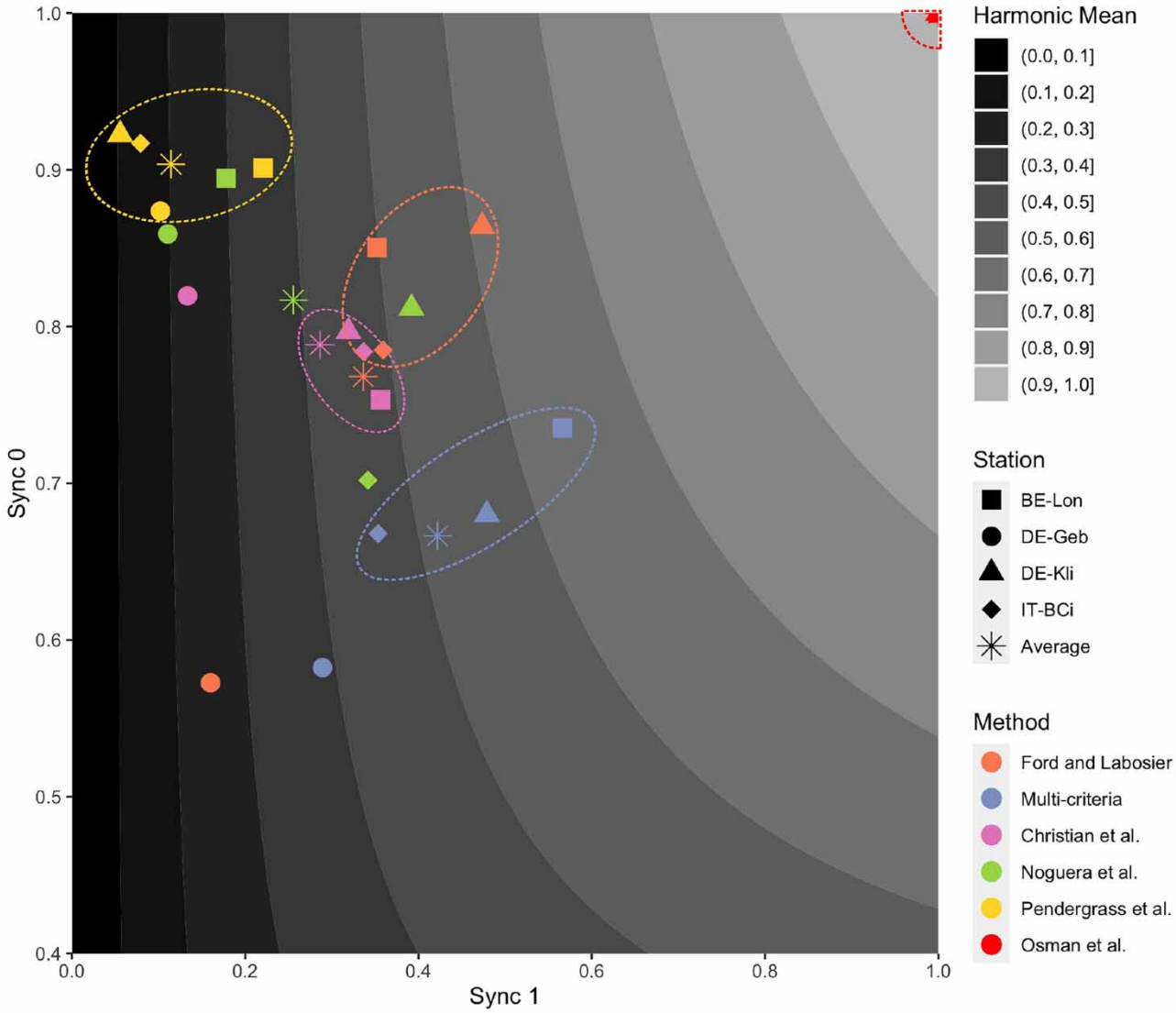


Figure 4 | Synchronicity metrics for the co-identification of FD events ($sync_1$) and intervals between FD events ($sync_0$) for Methods M2–M6, compared to M1 – Osman *et al.* as the reference method. The grey surface plot indicates the possible values of the harmonic mean $\overline{sync}$.

Noguera *et al.* and Pendergrass *et al.* had very low values, particularly for the TPR metric, with values below 0.3. Thus, method identification deviated significantly in two ways: the identification of multiple events that were not identified by the reference method and the omission of multiple events (e.g. M3 – Noguera *et al.* and M6 – Pendergrass *et al.*, both examples are also illustrated in Figure 3).

Both comparison metrics showed that even though there is a considerable resemblance in how certain methods identify FD events, there are just as many disparities. By studying each time series in Figure 4 in detail, four key factors were identified which explain the differences in FD detection as visualized in Figure 6:

A. Critical proxy variable opposed soil moisture dynamics

An example of this behaviour is shown in Figure 6(a), where monitored soil water content increased due to a longer rainfall period. However, M3 – Christian *et al.* identified an FD event using the SESR index, which increased considerably due to an ongoing increase in the potential ET rate.

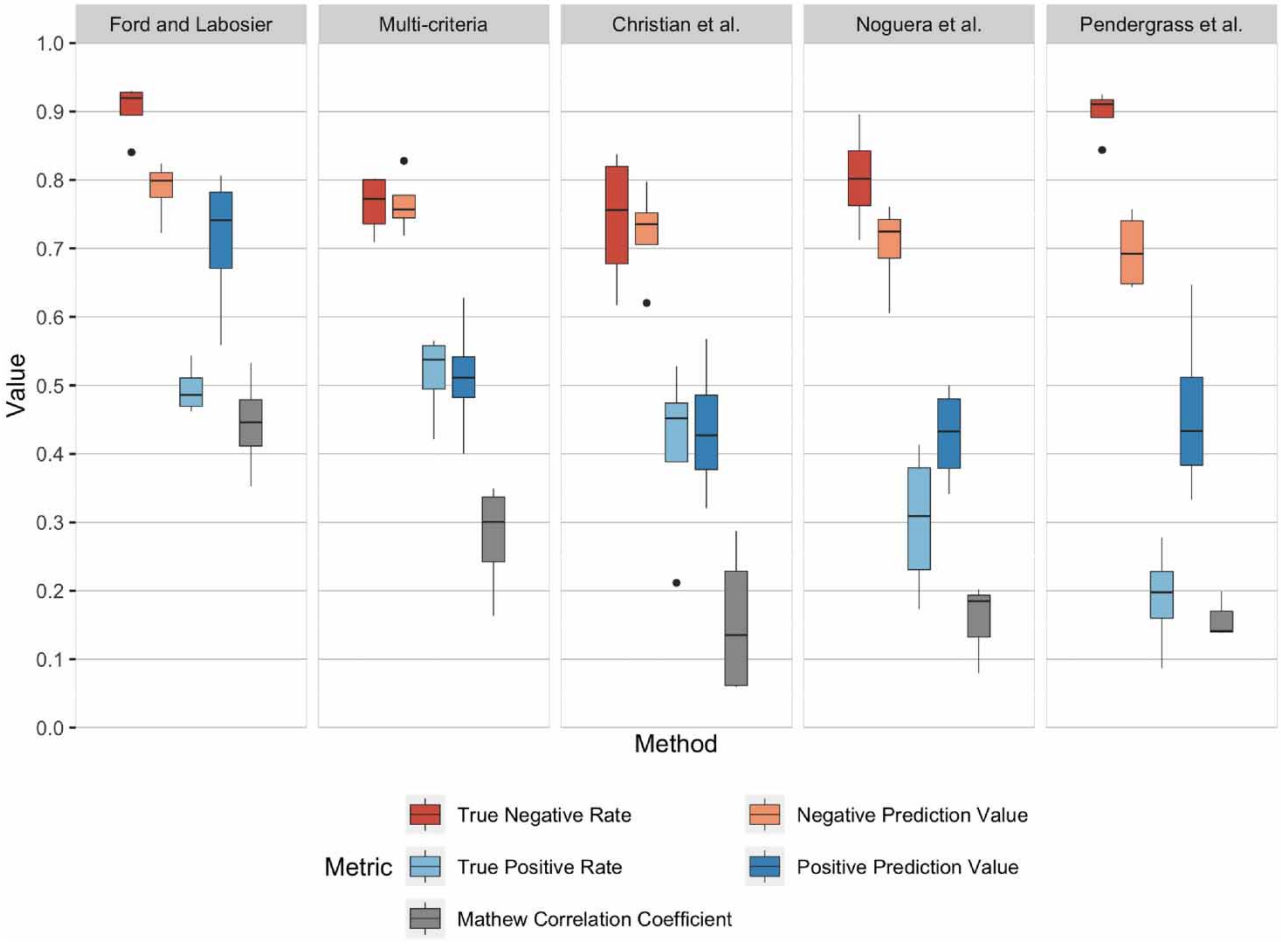


Figure 5 | Confusion matrix with multiple metrics evaluating true and false identifications of FD and no FD weeks of methods M2–M6 compared to those of M1 – Osman *et al.*, which was the reference method. Box plots represent all metrics including the data from all four stations and all weeks during the study period 2004–2014.

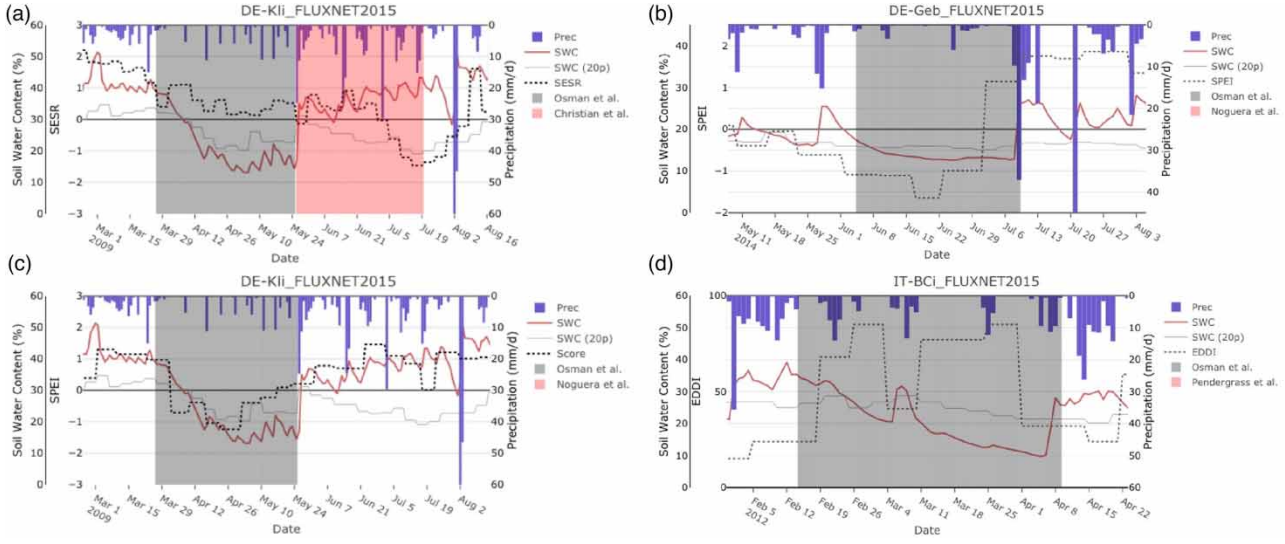


Figure 6 | Differences in FD detection due to four factors, illustrated with an example of misidentification: (a) critical proxy variable SESR of M3 – Christian *et al.* opposed moisture dynamics; (b) velocity of depletion was too small for M5 – Noguerra *et al.*; (c) threshold for the final depletion of M4: the multi-criteria score was not exceeded and (d) duration of the event was too short for M6 – Pendergrass *et al.*

B. Velocity of depletion was too small

As shown in Figure 6(b), the method M4 – Noguera *et al.* showed a decrease in their proxy variable SPEI during a period of rapid depletion of soil moisture. However, the time period over which the intensification occurred was not long enough to trigger an FD with M4 (but the reference method M1 was able to identify it).

C. Threshold for final depletion was not exceeded

In this case, the critical final threshold for FD identification is not exceeded. For example, the score of the M3 multi-criteria method did not surpass the set value of 0.65 during an actual incidence of rapid soil moisture loss (Figure 6(c)).

D. Duration of the event was too short

Finally, the event in Figure 6(d) shows an event missed by the EDDI of M6 – Pendergrass *et al.*, where a significant amount of precipitation caused a short recuperation, which did not break the course of rapid and extreme drying visible in the soil moisture series. However, the two intervals with a high EDDI were too short to be classified as FDs by M6.

Misidentification and near-misses (as illustrated in Figure 6) might have been due to just one of the factors but were often due to a combination of them.

3.4. Commonalities and disparities due to the lack of definition – the way forward: the ensemble approach

So far, we established that all six methods showed clear commonalities, but also disparities, in the detection and delineation of individual FD periods. The differences in the detected periods can be directly linked to the use of different critical variables and threshold values, and the different minimum durations and absolute or relative changes that are required for each method to detect an FD (Table 1).

We showed that the choice of the critical variable plays a major role in explaining the differences in identifications (Figure 6(a)). Different critical variables are already used in the analysis of 'normal' droughts, where the choice of variables such as rainfall, runoff and soil moisture deficits determines the type of drought form to be analysed. Normal droughts are commonly categorised into meteorological, hydrological and agricultural droughts (Mishra & Singh 2010). There are no comparable sub-definitions for FDs, although several studies have stated that some FDs are mainly influenced by heatwave dynamics and others state that they are more strongly influenced by rainfall deficits (Mo & Lettenmaier 2016; Pendergrass *et al.* 2020), but currently no clear definitions exist for such sub-types. FDs are known to be triggered or exacerbated by compound extreme climatic events (Otkin *et al.* 2018; Pendergrass *et al.* 2020; Lisonbee *et al.* 2021), but the relative magnitude and associated impact of each extreme have not yet been established (Mo & Lettenmaier 2016). At the same time, the methods differ in the extent to which they include compound dynamics in their detection procedure. Notably, the method M6 – Pendergrass *et al.* does not include any information on rainfall; therefore, it may be considered as an FD method that is solely driven by evaporative demand. Methods M3–M5 include a mixture of rainfall deficit, heatwave and ET information, whereas M1 and M2 use soil moisture as an integrative response variable for all vertical water exchange processes. We propose that by using multiple methods, one can identify different FD types, and future research efforts should concentrate on disentangling the extent to which different climatic components are responsible for the respective droughts.

The focus of this method comparison was placed on FDs in temperate croplands, where unusual rapid declining soil moisture during the growing season is considered the most direct indicator of severe plant water stress (Ford & Labosier 2017; Samaniego *et al.* 2018). However, methods using monitored, root-zone soil moisture data (M1 and M2) are severely limited by data restrictions, whereas indirect methods (M3–M6) using other proxy variables, such as rainfall or temperature, allow the use of much longer data series and more stations. Directly connected to the data limitation of the soil moisture-based methods are uncertainties in the derivation of the required percentiles. While they are impressive for soil moisture data, time series that are 11 or 14 years long, as the ones used here, are likely to result in skewed misrepresentation of the upper and lower percentiles required for the soil moisture-based methods. This fact highlights the dilemma of soil moisture-based approaches. While they should be preferred over other methods to correctly reproduce FD dynamics for croplands, the shortness of their series potentially leads to large uncertainties in their threshold values.

This study did not aim to determine whether some methods are better than others, but rather to verify whether the methods compare well, and thus evaluate whether any method could be used to assess FD dynamics for croplands in Central Europe. However, detailed synchronicity analysis does not confirm the latter. Similar to the comparison study of Osman *et al.* (2021) for the USA, we found different frequencies depending upon which definition was used. Nevertheless, three or more methods often detected the same FD (Figure 3). We propose that to balance the strengths and weaknesses of all methods, an ensemble

approach for event identification be used as a way forward; thereby not just one, but several methods may be employed, which can be easily implemented with our R Shiny app. Multiple co-identifications would thus diminish the uncertainty of incorrect identifications and might give a more comprehensive picture when different types of FDs occur (Wang & Yuan 2018).

Several questions remain unanswered. Although a rapid depletion in soil moisture was detected multiple times during the investigated growing periods, the impact on vegetation health remains unclear and needs to be assessed with information other than climate data (plant mortality or remotely sensed imagery, as done by Peters *et al.* (2002) and Otkin *et al.* (2019)). The impact of different soil types and their water-holding and storage capacities, as well as the impact of different crop and vegetation species on FD identification, require future assessment. Finally, a thorough long-term analysis of FD dynamics using an ensemble approach, as suggested here, is pending. However, the preliminary analysis showed that FDs on croplands in Central Europe are no rare events but occurred on average once every 1–2 years during the investigated time period.

4. CONCLUSIONS

Defining and delineating FDs pose new challenges for hydrologists and climatologists. In this study, we compared six different delineation methods on a local scale (climatological station) and observed a large degree of synchronicity, but also some divergence, in the identified FD periods depending on which definition was used. The disparities of one method for detecting different drought intervals in comparison to others were narrowed down to four factors: (1) the opposing behaviour of proxy climate variables in comparison to prevailing soil moisture dynamics; (2) differences in the estimated velocity of drought intensification; (3) not exceeding pre-set threshold values for final depletion; and/or (4) differences in pre-set minimum drought lengths. Rather than seeing the detected divergence in identifying drought periods among the various methods as a weakness, we suggest using an ensemble approach for event identification to ensure that FDs of different sub-types are detected. These include different combinations of compound drivers, with differences in intensification dynamics, and were validated in this study for the temperate central European region.

Compound events, such as compound warm spells and rainfall deficit, have become more frequent in the past decades (Zscheischler *et al.* 2020; Vogel *et al.* 2021) and are expected to increase. However, it is not clear to what degree the different components that are relevant for FD development will develop. Thus, a single, commonly accepted definition for FD delineation may be the wrong goal, as it would take away the flexibility that an ensemble approach using multiple methods has.

ACKNOWLEDGEMENTS

This research was supported by CAPES (Finance Code 001) and DAAD (Award No. 91693642). We would like to thank the authors of the flash drought methods, which we used in this study, for their attention and support during our implementation in R. They include Dr Trent Ford, Dr Angeline Pendergrass, Dr Jordan Christian, Dr Mahmoud Osman, Dr Mike Hobbins and Ivan Noguera.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Allen, R. G. 1998 *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. Food and Agriculture Organization of the United Nations, Rome. Available from: http://www.fao.org/3/x0490e/x0490e01.htm#preface.

Chen, L. G., Gottschalck, J., Hartman, A., Miskus, D., Tinker, R. & Artusa, A. 2019 Flash drought characteristics based on U.S. drought monitor. *Atmosphere* **10** (9), 498. doi:10.3390/atmos10090498.

Chicco, D., Tötsch, N. & Jurman, G. 2021 The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* **14** (1). doi:10.1186/s13040-021-00244-z.

Christian, J. I., Basara, J. B., Otkin, J. A. & Hunt, E. D. 2019a Regional characteristics of flash droughts across the United States. *Environmental Research Communications* **1** (12), 125004. doi:10.1088/2515-7620/ab50ca.

Christian, J. I., Basara, J. B., Otkin, J. A., Hunt, E. D., Wakefield, R. A., Flanagan, P. X. & Xiao, X. 2019b A methodology for flash drought identification: application of flash drought frequency across the United States. *Journal of Hydrometeorology* **20** (5), 833–846. doi:10.1175/jhm-d-18-0198.1.

Christian, J. I., Basara, J. B., Hunt, E. D., Otkin, J. A. & Xiao, X. 2020 Flash drought development and cascading impacts associated with the 2010 Russian heatwave. *Environmental Research Letters* **15** (9), 094078. doi:10.1088/1748-9326/ab9faf.

de Araújo, J. C. & Bronstert, A. 2015 A method to assess hydrological drought in semi-arid environments and its application to the Jaguaribe River Basin, Brazil. *Water International* **41** (2), 213–230. doi:10.1080/ 02508060.2015.1113077.

Delgado, R. & Tibau, X.-A. 2019 Why Cohen's kappa should be avoided as performance measure in classification. *PLoS ONE* **14** (9), e0222916. doi:10.1371/ journal.pone.0222916.

D'Odorico, P., Davis, K. F., Rosa, L., Carr, J. A., Chiarelli, D., Dell'Angelo, J. & Rulli, M. C. 2018 The global food-energy-water nexus. *Reviews of Geophysics* **56** (3), 456–531. doi:10.1029/2017rg000591.

FAO 2021 *The Impact of Disasters and Crises on Agriculture and Food Security: 2021*. FAO. doi:10.4060/cb3673en.

Ford, T. W. & Labosier, C. F. 2017 Meteorological conditions associated with the onset of flash drought in the eastern United States. *Agricultural and Forest Meteorology* **247**, 414–423. doi:10.1016/j.agrformet.2017.08.031.

Hobbins, M. T., Wood, A., McEvoy, D. J., Huntington, J. L., Morton, C., Anderson, M. & Hain, C. 2016 The evaporative demand drought index. Part I: linking drought evolution to variations in evaporative demand. *Journal of Hydrometeorology* **17** (6), 1745–1761. doi:10.1175/jhm-d-15-0121.1.

James, G., Witten, D., Hastie, T. & Tibshirani, R. 2013 *An Introduction to Statistical Learning*, Vol. 112. Springer, New York.

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S. & Blöschl, G. 2020 Joint trends in flood magnitudes and spatial extents across Europe. *Geophysical Research Letters* **47** (7). doi:10.1029/2020gl087464.

Lesk, C., Rowhani, P. & Ramankutty, N. 2016 Influence of extreme weather disasters on global crop production. *Nature* **529** (7584), 84–87. doi:10.1038/ nature16467.

Liang, M. & Yuan, X. 2021 Critical role of soil moisture memory in predicting 2012 central USA flash drought. *Frontiers in Earth Science* **9**, 46.

Lisonbee, J., Woloszyn, M. & Skumanich, M. 2021 Making sense of flash drought: definitions, indicators, and where we go from here. *Journal of Applied and Service Climatology* **2021** (1), 1–19. doi:10.46275/joasc.2021.02.001.

Lukas, J., Hobbins, M. & Rangwala, I. 2017 *The Eddi User Guide v1.0*. National Integrated Drought Information System (NIDIS). Available from: https:// psl.noaa.gov/eddi/.

Matthews, B. 1975 Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) – Protein Structure* **405** (2), 442–451. doi:10.1016/0005-2795(75)90109-9.

Mishra, A. K. & Singh, V. P. 2010 A review of drought concepts. *Journal of Hydrology* **391** (1–2), 202–216. http://dx.doi.org/10.1016/j.jhydrol.2010.07.012.

Mo, K. C. & Lettenmaier, D. P. 2015 Heat wave flash droughts in decline. *Geophysical Research Letters* **42** (8), 2823–2829. https://doi.org/10.1002/2015GL064018.

Mo, K. C. & Lettenmaier, D. P. 2016 Precipitation deficit flash droughts over the United States. *Journal of Hydrometeorology* **17** (4), 1169–1184. doi:10 .1175/jhm-d-15-0158.1.

Nguyen, H., Wheeler, M. C., Hendon, H. H., Lim, E.-P. & Otkin, J. A. 2021 The 2019 flash droughts in subtropical eastern Australia and their association with large-scale climate drivers. *Weather and Climate Extremes* **32**, 100321. doi:10.1016/j.wace.2021.100321.

Noguera, I., Domínguez-Castro, F. & Vicente-Serrano, S. M. 2020 Characteristics and trends of flash droughts in Spain, 1961–2018. *Annals of the New York Academy of Sciences* **1472** (1), 155–172. https://doi.org/10.1111/nyas.14365.

Oikonomou, P. D., Karavitis, C. A., Tsesmelis, D. E., Kolokytha, E. & Maia, R. 2020 Drought characteristics assessment in Europe over the past 50 years. *Water Resources Management* **34** (15), 4757–4772. doi:10.1007/ s11269-020-02688-0.

Osman, M., Zaitchik, B. F., Badr, H. S., Christian, J. I., Tadesse, T., Otkin, J. A. & Anderson, M. C. 2021 Flash drought onset over the contiguous United States: sensitivity of inventories and trends to quantitative definitions. *Hydrology and Earth System Sciences* **25** (2), 565–581. doi:10.5194/hess-25-565-2021.

Otkin, J. A., Anderson, M. C., Hain, C., Mladenova, I. E., Basara, J. B. & Svoboda, M. 2013 Examining rapid onset drought development using the thermal infrared-based evaporative stress index. *Journal of Hydrometeorology* **14** (4), 1057–1074. doi:10.1175/jhm-d-12-0144.1.

Otkin, J. A., Svoboda, M., Hunt, E. D., Ford, T. W., Anderson, M. C., Hain, C. & Basara, J. B. 2018 Flash droughts: a review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bulletin of the American Meteorological Society* **99** (5), 911–919. doi:10.1175/bams-d-17-0149.1.

Otkin, J. A., Zhong, Y., Hunt, E. D., Basara, J., Svoboda, M., Anderson, M. C. & Hain, C. 2019 Assessing the evolution of soil moisture and vegetation conditions during a flash drought–flash recovery sequence over the southcentral United States. *Journal of Hydrometeorology* **20** (3), 549–562. doi:10.1175/jhm-d-18-0171.1.

Palmer, W. C. 1965 *Meteorological Drought*, Vol. 30. U.S. Department of Commerce, Weather Bureau, Washington DC.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W. & Papale, D. 2020 The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data* **7** (1), 225. https://doi.org/10.1038/s41597-020-0534-3.

Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A. & Woodhouse, C. A. 2020 Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nature Climate Change* **10** (3), 191–199. doi:10.1038/s41558-020-0709-0.

Peters, A. J., Walter-Shea, E. A., Ji, L., Vina, A., Hayes, M. & Svoboda, M. D. 2002 Drought monitoring with NDVI-based standardized vegetation index. *Photogrammetric Engineering and Remote Sensing* **68** (1), 71–75.

Rogers, R. R. 1989 *A Short Course in Cloud Physics*. Butterworth-Heinemann, Oxford, UK/Boston, MA.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M. & Marx, A. 2018 Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change* **8** (5), 421–426. doi:10.1038/ s41558-018-0138-5.

Vogel, J., Paton, E., Aich, V. & Bronstert, A. 2021 Increasing compound warm spells and droughts in the Mediterranean Basin. *Weather and Climate Extremes* **32**, 100312. doi:10.1016/j.wace.2021.100312.

Wang, L. & Yuan, X. 2018 Two types of flash drought and their connections with seasonal drought. *Advances in Atmospheric Sciences* **35** (12), 1478–1490. doi:10.1007/s00376-018-8047-0.

Yuan, X., Wang, L., Wu, P., Ji, P., Sheffield, J. & Zhang, M. 2019 Anthropogenic shift towards higher risk of flash drought over China. *Nature Communications* **10** (1). doi:10.138/s41467-019-12692-7.

Zscheischler, J., van den Hurk, B., Ward, P. J., Westra, S., 2020 Multivariate extremes and compound events. In: *Climate Extremes and Their Implications for Impact and Risk Assessment* (Sillmann, J., Sippel, S. & Russo, S., eds). Elsevier, Amsterdam. doi:10.1016/B978-0-12-814895-2.00004-5.

First received 21 December 2021; accepted in revised form 25 May 2022. Available online 8 August 2022